Attorney Docket No. 50269-0026

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BOARD OF PATENT APPEALS AND INTERFERENCES

#18
10-21-03

In re application of	Confirmation No.: 3744
Michael E. Palmer et al.	Group Art Unit No.: 2176
Serial No.: 09/333,121	Examiner: Charles A. Bieneman
Filed on: June 14, 1999	
For: METHOD AND APPARATUS FOR MEASURING SIMILARITY AMONG ELECTRONIC DOCUMENTS	

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

Sir:

This is an appeal from the Examiner's rejection in the Final Office Action mailed May 23, 2003. This Appeal Brief is submitted in support of the Notice of Appeal filed August 15, 2003, and received August 18, 2003. Since the October 18, 2003 was a Saturday, the period for response extends until the following Monday, which is October 20, 2003.

10/24/2003 HDEMESS1 00000081 09333121

01 FC:1402 330.00 OP

I. REAL PARTY IN INTEREST

Yahoo!, of Sunnyvale, California (this application was formerly owned by Inktomi Corporation, but Inktomi Corporation and its intellectual property were assigned Yahoo!).

II. RELATED APPEALS AND INTERFERENCES

None that the Applicant is aware of.

III. STATUS OF CLAIMS

Claims 1-34 are pending.

Claims 1 –14 , 17 –19, and 26 – 34 stand rejected under 35 U.S.C. § 103(a) as being allegedly upatentable over U.S. Patent No. 5,835,905, herein *Pirolli*, in view of U.S. Patent No. 5,690,422, herein *Prasad*.

Claims 15 and 16 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Pirolli* and *Prasad* "as applied to claim 1" and further in view of U.S. Patent No. 6,282,549, herein *Hoffert*.

Claim 20 stands rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Pirolli* and *Prasad* "as applied to claim 1" and further in view of U.S. Patent No. 6,128,606, herein *Bengio*.

Claims 21-25 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over *Pirolli* and *Prasad* "as applied to claim 1" and further in view of U.S. Patent No. 6,389,436, herein *Chakrabarti*.

IV. STATUS OF AMENDMENTS

An amendment After Final was filed on June 25, 2003, but it was not entered.

A second response after final was filed on August 15, 2003, but it did not include any amendments. No amendments to the specification or the claims have been made after the final reply to the first Office Action dated August 15, 2003.

V. SUMMARY OF THE INVENTION

Briefly a training set of documents is established having a group of documents in each category. A similarity matrix is established and may be used to form an objective function that may be optimized. The optimization process is performed by adjusting scores associated with documents that represent the strength or degree that a document belongs to a category.

FIG. 1 is a block diagram depicting various sources of similarity information that are fed to a Similarity Objective Function 110, which is the objective function that is optimized during the classification process. The measures of document similarity may include, but are not limited to, hyperlink similarity (taken from hyperlink info 100), the similarity of the text of the documents (taken form text similarity 102), multimedia similarity (taken from multimedia component similarity 104), URL similarity 106, and user click-through similarity (taken from user click info 108), for example. These measures of similarity are discussed in turn below.

Regarding hyperlink info 100, weights are assigned corresponding to hyperlinks from one document to another. Some hyperlinks have greater or lesser similarity weight than other hyperlinks, based on other features of the links or their source or destination documents.

Regarding text similarity 102, two documents may be considered similar based on a comparison of word vectors derived from the text of each of the two documents. Text similarity may be determined in part based upon weight values assigned to words of the text. Some words may be given greater or lesser weight than other words.

Regarding multimedia component similarity 104, two documents may be considered to have multimedia similarity based on the similarity of features derived from multimedia components linked to or contained by the documents. Regarding URL similarity 106, two documents may be considered similar if a URL of each document contains similar URL sub-components.

Regarding click through similarity 108, two documents are considered similar based on user click-through similarity when the documents are associated with similar patterns of user click behavior. Some behaviors from among which click through similarity may be based are the frequency of clicks, the click context, the duration of viewing, the proximity in time to other clicks, and the proximity in context to other clicks. The measures of document similarity may be derived from patterns detected in user viewing of the documents, and the user viewing information may be monitored by a web caching system and stored in a log. Some patterns of viewing behavior that may used to determine document similarity are frequency of viewing, viewing context, duration of viewing, proximity in time to other documents viewed by the same user, or similarity of patterns of viewing by all users.

Each of the sources of similarity information 100, 102, 104, 106, and 108, which are represented as a graph of links, is fed into a combination function 130 to produce a combined graph 140. The combination of two or more measures of document similarity (e.g., 100, 102, 104, 106, or 108) may be achieved by taking the union of each of a plurality of graphs in

which each graph describes one of the measures of document similarity. Based on the union of the two graphs, a combined graph may be computed that describes the combined document similarity. The combination of two or more measures of document similarity may also be computed by taking the intersection of each of a plurality of graphs (in which each graph describes one of the measures of document similarity) to derive a combined graph (that describes the combined document similarity). The combined similarity graph 130 and the similarity objective function 110 are used to compute a generalized similarity value 120 for two exemplary documents 112 and 114 that are stored in a hypertext system. The similarity information may be extracted from the similarity matrix to obtain new documents that are supported by the set of training documents for each category, and therefore categorized within the corresponding one or more categories.

FIGs. 3A and 3B generally show an example of the classification process. FIGs. 3 A and 3B and the corresponding example are first described. The classification process is then summarized in more general terms, and further details are given.

FIG. 3A is a flow diagram of a method of computing a Similarity value using a global similarity objective function. In the specification, for purposes of clarity, implementation of the method is illustrated by an example in which only text similarity and hyperlink similarity are used, and a gradient search algorithm is employed.

The Data Preparation step 302 is described in conjunction with FIG. 3B.

Next, the Pre-processing step 304 calculates one or more Similarity Matrices for different types of similarity. These matrices are combined using the combination function to calculate the Combined Similarity Matrix.

Subsequently, the Generalized Similarity Training step 306 takes documents in the training set of each category, and finds documents that are similar to them by minimizing an objective function. The documents are thereby classified into the corresponding categories.

Next, the Post-processing step 308 may generally perform a clean up of the results by heuristic methods that have been found in practice to improve results somewhat. For example, Post-processing step 308 may involve removal of documents that have been determined to be "spam" documents by a wide variety of heuristic methods.

FIG. 3B is a block diagram of steps that may be involved in an embodiment of Data Preparation step 302. In block 310, a training set is created. The training set may include a small set of electronic documents that are determined to closely match each category of a taxonomy. Each document may belong to more than one category and may be marked to indicate the categories to which it belongs. Next, data used in Generalized Similarity Training 306 is generated. A graph may be extracted from the documents and constructed in memory based on an expanded set of documents. In this example, text similarity and link similarity are used. The expanded set of documents is created by expanding the training set to include all the documents that the training documents point to or that point to the training set documents. A link graph is created and stored to represent the link relationship among all documents in the expanded set, as indicated by block 312.

In block 314, using the text contents of each document, possible single- and multiple-word phrases are extracted from the documents. Feature analysis and extraction is also carried out on the documents. A subset of the features that most strongly discriminate documents in one category from documents in another category are selected. As shown in block 316, word vectors or feature vectors are constructed for each document in the expanded

set. Each component of the feature vectors is the normalized value of the occurrence frequency of a particular feature in this document. (The components of these vectors are adjusted during generalized similarity training 306 to classify the documents.)

To reiterate and further elaborate, as disclosed, a plurality of new electronic documents are categorized into categories by establishing a plurality of training sets. Each training set is associated with a category and includes training documents that have been classified as belonging to that category. A determination is made regarding how strongly each document of the plurality of documents corresponds to each of said plurality of categories by determining similarity between each of the documents and the training documents that belong to the training set of the categories. The determination of the similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity. An objective function referencing attributes associated with each of said plurality of documents including the matrix is optimized (e.g., maximized, minimize, or otherwise extremized, as generally discussed starting at page 20, line 6, and ending on page 29, line 3).

Alternatively, the similarity information may be obtained by only approximately optimizing the objective function. Various transformations are used to perform the relaxation process used in the optimization. Examples of these transformations are given by equations II-5, II-7, and II-8, and the equation of page 28, lines 14-16. The training process involves optimizing (e.g., maximizing) an objective function P(x). Examples of the objective function are given in equations II-1 and II-9, and the equation of page 21, line 21. The i's and j's in these equations (and in the similarity matrix W_{ij}) are indices that correspond to documents (see page 22, lines 9-11, for example). The similarity matrix W_{ij} represents the similarity

between documents, and is formed from feature vectors S(i,k) (see the equation on page 13, line 23, for example). The feature vectors are based on various matrices used to characterize the attributes of the documents (page 13, lines 9-13 give an example of a feature vector).

Thus, an optimization (recited in the specification) is performed in the training process, which calculates a maximum or optimum value for the objective function (as explained on page 23, lines 5-8, cited above). The similarity matrix W_{ij} is included in the objective function in equations II-1 and II-9, and the equation of page 21, line 21. Confidence scores x_i or $x_{i\beta}$ are associated with documents that are referenced by the objective equation (or function) via their presence in the equation used and are attributes that reference the document they are used to rank. In addition to the indices of the similarity matrix W_{ij} referring to documents, each element of the matrix W_{ij} is derived via the feature vectors from features or attributes of the document, and in this manner the objective function references attributes associated the documents to which indices *ij* refer.

The optimization of the objective function may be performed by repeated application of a growth transformation. During the optimization process a second matrix may be created and stored that represents an interim score for each document in each category. The second matrix may be created and stored using columns (for example) to represent documents and rows (for example) to represent user sessions. Alternatively, the rows may represent documents and the columns may represent users. The values of elements of the second matrix may represent interest in a document shown by a particular user in a particular session. The element values may be a function of the time that a user has spent viewing a document associated with each element. The second matrix may represent a similarity between pairs of documents represented by indices labeled i and j, for example. The second matrix may be

derived by comparing pairs of column vectors or row vectors, respectively, of the i and j columns or rows of the first matrix.

Periodically, as the similarity matrix is being computed, the rows of the similarity matrix may be normalized. The normalization may be performed by normalizing the representation of each document (e.g., the scores), and the normalization may be performed across all categories. The normalization may be performed in such a way that the score for one document in a particular category will depend on the scores for that document in all other categories. Similarly, the columns of the similarity matrix may be normalized by performing a normalization within each category, across all documents in such a way that the score for one document in a particular category depends on the scores for all other documents in that category. For example, a similarity between pairs of documents i and j, may be determined by finding pairs of documents i and j that have high interest values for a particular user in a particular session or period of time.

The categories in to which the training set is placed may come from a manually defined taxonomy or derived from logs of user queries, for example. The categories may be derived by identifying a category of a classification taxonomy of a hypertext system in which a first electronic document is presently classified. If a second electronic document is found to be highly similar, information may be stored that classifies the second electronic document into the category.

VI. ISSUES

A. Claims 1 and 34 are not obvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

B. Claims 8 is not obvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

C. Claims 9 is not obvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

D. Claims 17 and 18 are not obvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

E. Claim 18 is not obvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

F. Claim 20 is not obvious, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of U.S. Patent No. 6,128,606, herein *Bengio.*

G. Claim 20 is not obvious, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of *Bengio*, and claims 21-25 are not obvious, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of *Chakrabarti.*

VII. GROUPING OF CLAIMS

None of the claims should be regarded as all standing together since each claim recites limitations that render it separately patentable. However, for the purposes of this appeal, the following groups are recognized:

Claims 1-16, 19, 21-23, and 26-34 (not argued separately)

Claims 8 and 9 (each argued separately and do not stand and fall together)

Claims 17 and 18 (each argued separately and do not stand and fall together)

Claim 20

Claims 21 and 23-25 (not argued separately)

Claim 22

VIII. ARGUMENTS

A. OUTLINE OF ARGUMENTS

The arguments are separated into the following hierarchical headings.

A. Outline of Arguments

B. Claims 1 and 34 are not Obvious Over Pirolli in view of Prasad (issues A)

(1) Reference Combination Does Not Meet All Features Recited in the Claims

(a) The Combination of Pirolli and Prasad Lacks a Teaching of Comparing a Group of Documents to a Group of Documents

(i) Pirolli Compares Documents to Rules and Not to Documents

(ii) Prasad also Uses Rules for Directing Queries Rather than Making a Comparison to the Documents Themselves to Determine Categories

(iii) Combination of Pirolli and Prasad Would Still Use a Comparison of Documents to Rules Rather Than Documents to Documents

(b) Activation of Pirolli is not a Sub-step of Categorizing

(3) Combination of Pirolli and Prasad is Improper in a Rejection Under 35 USC 103

(a) Prasad Teaches Away From The Claimed Categorization by a Comparison to Documents

(b) Pirolli Teaches Away From the Claimed Invention

(i) Categories in Pirolli Either do Not Rely on Similarity or Do Not have Documents Established as Belonging to Them

(ii) Differences in Types of Categories Mitigate against Combining Pirolli and Prasad

(iii) Pirolli Teaches Global Rules, Requiring Human Intervention to Arrive at the Rules, Discourage Prasad's Automated Comparison to Documents to Arrive at Rules

(iv) Added Effort in Applying Method of Prasad to Priolli as Compared to Applying Prasad in General is a Further Deterrent to Combining Prasad and Pirolli

(c) With Hindsight Removed, One Would not Have Arrived at the Modification Proposed by the Examiner

(d) Prasad and Pirolli are from different fields of endeavor

C. Claim 8 is Not Obvious Because the Examiner has Not Shown Click Through Behavior (issue B)

D. Claim 9 is Not Obvious Because the Examiner has Not Shown Click Through Behavior (issue C)

E. Claims 17 and 18 are Not Obvious Because the Examiner has Not Shown Forming a Graphs From Two Graphs (issue D)

F. Claim 18 is Not Obvious Because the Examiner has Not Shown an Intersection of Two Graphs (issue E)

G. Claim 20 is Not Obvious Over Bengio Because the Examiner has Not Provided a Proper Motivation to Combine and Bengio is Nonanalogous Art (issue F)

H. Claims 20-25 are not Obvious Because Pumping Activation is Not an Optimization of an Objective Function (issue G)

I. Remaining Dependent Claims

B. CLAIMS 1 AND 34 ARE NOT OBVIOUS OVER PIROLLI IN VIEW OF PRASAD

The Examiner rejected claims 1 –14, 17 –19, and 26 – 34 are nonobvious, under 35 U.S.C. § 103(a), over *Pirolli* in view of *Prasad.*

(1) REFERENCE COMBINATION DOES NOT MEET ALL FEATURES RECITED IN THE CLAIMS

As stated in MPEP 2143.03, (entitled, "All Claim Limitations Must Be Taught or Suggested"), "To establish *prima facie* obviousness of a claimed invention, all the claim

limitations must be taught or suggested by the prior art. *In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)." Independent claims 1 and 34 contain features not shown in the combination of *Pirolli* and *Prasad*, as discussed below.

(a) ***The Combination of Pirolli and Prasad Lacks a Teaching of Comparing a Group of Documents to a Group of Documents***

Below, first in subsection "(i)" it is shown that a comparison of documents to documents is not shown in Pirolli. Then in subsection "(ii)" it is shown that a comparison of documents to documents is not suggested in Prasad. Finally, in subsection "(iii)" it is shown that the combination of Pirolli and Prasad also do not show the claimed comparison of documents to documents.

(i) *Pirolli Compares Documents to Rules and Not to Documents*

Claims 1 and 34 recite

> establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;
> determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories and the documents that belong to the training set of said category ...

This passage of claims 1 and 34 requires that a comparison of a group of documents to a group of documents is used in categorizing documents. To elaborate, the Applicant admits that *Pirolli* teaches (1) to categorize a set of documents, in the form of pages, according to "classification characteristics", and (2) to determine textual similarity between documents in order to categorize a document. However, Applicant is not attempting to claim only these features. Rather, the Applicant is claiming the use of the similarity between a plurality of

documents and particular sets of documents (i.e., a training set), which have been established as belonging to a category, to determine the correspondence between the document and the category. However, in contrast to claims 1 and 34, *Pirolli* classifies by comparing to a rule rather than to a plurality of documents. For example, *Pirolli* state, "The classification characteristics are predetermined 'rules' ...." (column 5, lines 13 an14). Similarly, *Pirolli* state, "The present invention utilizes an approach based on weighted linear equations that define the ***rules*** for predicting degree of category membership..." (emphasis added, column 8, lines 41 and 42).

(ii) *Prasad Also Uses Rules for Directing Queries Rather Than Making a Comparison to the Documents Themselves to Determine Categories*

Similar to *Pirolli*, *Prasad* also fails to teach the claimed feature of using similarity between a document and another set of documents established as belonging to a category to determine the correspondence between the document and the category. Presumably, the Examiner has equated a document as claimed to a document at a data source and a training set as claimed to a sample of documents from a data source. Even if the training set taught by *Prasad* can be equated to the training set claimed, *Prasad* nevertheless fails to teach the claimed feature of comparing a document to documents in a training set. Specifically, *Prasad* teaches (at col. 3, line 66 – col. 4, line 16)

> In FIG. 2, a plurality of data sources $20^1 \ldots 20^n$, each source containing a plurality of documents $20^1 \ldots 20^n$ are available for searching in response to a query entered into the system 10 by a user. The data sources are stored in the databases 18 associated with the servers 14 (See FIG. 1). As a solution to providing an automatic and optimal selection of desired data sources for user queries, a form of supervised machine learning called "Rule Induction" generates a model for classifying the sources 20 for query searching. The model is then used for predicting the top "N" sources most likely to contain documents that satisfy a user's query. As an overview,

> ***"Rule Induction" takes a sample set of documents called a training set and derives "Disjunctive Normal Form Rules" representative of the model*** which is descriptive of the data sources 20. "Rule Induction" is often the preferred approach to classification modeling and prediction due to the enhanced capability and interpretability of decision rules in responding to user queries (emphasis added).

In other words, *Prasad* teaches that rule induction is applied to the training set to first generate rules, and then the rules (and not the training documents) are used to determine what source to direct queries. In other words, the rules are used to direct search queries and not to categorize documents, and even if *arguendo* directing search queries were a categorization of documents, a comparison to "rules" is used for the "directing" and a comparison to documents is not used for the directing.

Regarding the Applicants reliance on the above passage (col. 3, line 66 – col. 4, line 16), the Examiner stated (paper #11, page 4, the second paragraph),

> The examiner disagrees with applicants' characterization of *Prasad* inasmuch as the rule induction taught by *Prasad* is used to classify documents, i.e., determine their similarity to a category. (*Prasad*, col. 4, lines 3-16.)

The Applicants respectfully submit that contrary to the implications of the Final Office Action, column 4, lines 3-16, of *Prasad* never suggests that "the rule induction taught by *Prasad* is used to classify documents, i.e., determine their similarity to a category." Instead, column 4, lines 3-16 state, " 'Rule Induction' generates a model for classifying the sources **20** for query searching." In other words, the rules derived by rule induction are used to direct queries not to classify documents. The interpretation of column 4, lines 3-16, as referring to (1) using rules rather than a comparison to documents and (2) using the rules to characterize sources of documents and not the documents is supported by other passages of Prasad. Specifically, the generation of the model is performed by (column 3, lines 19-23)

> A prior art algorithm is used to recognize patterns in the sets of samples to distinguish one ***source*** from another and generate a set of Disjunctive Normal Form (DNF) ***Rules***, as a model, representing each ***source***. (emphasis added)

Alternatively, as stated in column 4, lines 10-13,

> "Rule Induction" takes a sample set of documents called a training set and dervies "Disjunctive Normal Form Rules" representative of the model which is descriptive of the data ***sources*** 20... (emphasis added)

In other words, the sources 20 are the "categories" into which the documents are already located, and in this sense preclassified, and rules are derived for determining the common characteristics of the documents that distinguish them from the documents of other sources. (However, in Prasad, no new documents are classified into these categories, because the categories are the sources for finding the sources.) For example (column 3, lines 16-19),

> A dictionary is created to define features and attributes representing individual ***sources***. All documents are transformed into a set of samples comprising a feature, a word or phrase and a source name used in the dictionary. (emphasis added)

After deriving rules for the sources (column 4, lines 7-9),

> The model is then used for predicting the top "N" sources most likely to contain documents that satisfy a user's query.

Thus, in view of the above passages (which are column 3, lines 16-19 and 19-23 and column 4, lines 7-9), it can be seen that column 4, lines 3-16, cited by the Examiner, disclose using documents in a source to derive characteristics of a source for formulating rules that are used for finding which source is most likely to contain a document that meets a search query. Column 4, lines 3-16, also do not disclose classifying new documents by comparing them to other documents. Instead, *Prasad* teaches that rule induction is applied to the training set to generate rules that are used to determine which source to direct queries. While *Prasad*

teaches that training sets are used as input for rule induction, no teaching in *Prasad* suggests training sets themselves are used determine the correspondence between a document (or even a search query) and the category to which the training set belongs by determining the similarity between the document (or even a search query) and the actual training set.

(iii) *Combination of Pirolli and Prasad Would Still Use a Comparison of Documents to Rules Rather Than Documents to Documents*

Thus, contrary to the Examiner's assertions (in paper # 11 - the Final Office Action - at the bottom of page 3), as shown above, both Pirolli and Prasad use a comparison of rules to documents rather than documents to documents for categorizing documents or dierting queries. Pirolli uses rules for predicting the degree of category membership. Prasad uses rules, established via a training set, to direct queries. Since both *Prasad* and *Pirolli* use rules rather than a comparison to documents to categorize documents or direct queries, the combination of the two references also cannot suggest comparing a plurality of documents to sets of documents to categorize the documents.

(b) *Activation of Pirolli is Not a Sub-step of Categorizing*

The Examiner (in paper # 9 at the bottom of page 3) relied on the *Pirolli*'s teaching of establishing categories for meeting the limitation of "determining how strongly each of the documents corresponds to each of said categories" (in the second to last paragraph of claims 1 and 34, cited above). The Examiner (in paper # 9 at the first full paragraph of page 4) also relied upon *Pirolli*'s teaching of spreading activation to meet the limitation of "determining similarity" (recited in the next line of claims 1 and 34), and which also is performed using a

matrix (as recited in the next paragraph of claims 1 and 34). Further, the Examiner needs to make these associations in this manner so that the matrix used for spreading activation is the matrix associated with the step of determining similarity (as recited in the "wherein" clause). Referring to FIG. 1 of *Pirolli*, the classification is provided (step 103) and applied to feature vectors (step 104) as a preparation for performing the spreading activation (step 106). In contrast, the above excerpt of claims 1 and 34 recites, "determining how strongly each document ... corresponds to each of said plurality of categories [which the Examiner associated with categorization of the training set] ***by*** determining similarity [which the Examiner associated with spreading activation]" (emphasis added). Thus, claims 1 and 34 require the "determining similarity..." (and therefore according to the Examiner the spreading activation) to be a sub-step of the step of "determining how strongly..." (and therefore to be a sub-step of the categorization of the training set, following the Examiner's line of reasoning), and in contrast, in *Pirolli* the categorization steps are preparation for the spreading activation and not performed by the spreading activation. Thus, again the Examiner has not met his burden of proof, because the Examiner's explanation is inconsistent regarding the claim recitation of "determining how strongly each document ... corresponds to each of said plurality of categories ***by*** determining similarity... wherein the ... determining of similarity is performed using a matrix ... that is derived by combining two or more measures of document similarity."

(3) COMBINATION OF PIROLLI AND PRASAD IS IMPROPER IN A REJECTION UNDER 35 USC 103

In order for a combination of references to be proper under 35 USC 103, there must be a motivation to combine the references in a manner that results in the claimed invention. For example, MPEP 2143.01, (under the title "THE PRIOR ART MUST SUGGEST THE DESIRABILITY OF THE CLAIMED INVENTION") states, "Obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so.... *In re Kotzab,* 217 F.3d 1365, 1370, 55 USPQ2d 1313, 1317 (Fed. Cir. 2000). See also >*In re Lee,* 277 F.3d 1338, 1342-44, 61 USPQ2d 1430, 1433-34 (Fed. Cir. 2002) (discussing the importance of relying on objective evidence and making specific factual findings with respect to the motivation to combine references);< *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988); *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992)." Although the Examiner often attempts to provide a motivation, the motivations provided by the Examiner are deficient for the reason explained below. For example, logically if one of ordinary skill would expect that the proposed motivation would in fact provide no benefit or the rationale for the proposed motivation is inapplicable to the combination actually being proposed, it logically follows that there is in fact no motivation.

(a) ***Prasad Teaches Away From the Claimed Categorization by a Comparison to Documents***

One deficiency in a motivation to combine references is when the references also provide a teaching that teaches away from making the proposed modification. As stated in

MPEP 2144.05, p. 2100-138, "A *prima facie* case of obviousness may also be rebutted by showing that the art, in any material respect, teaches away from the claimed invention. *In re Geisler*, 116 F.3d 1465, 1471, 43 USPQ2d 1362, 1366 (Fed. Cir. 1997)." Logically, one of ordinary skill in the art would ignore what might otherwise be interpreted as a motivation to combine, if there is a suggestion that in the case at hand the combination is undesirable.

Prasad states "'Rule Induction' is often the preferred approach to classification modeling and prediction due to the enhanced capability and interpretability of decision rules in responding to user queries" (column 4, lines 3-16, cited above). Thus, Prasad prefers rules for directing queries rather than a direct comparison to documents because of their enhanced capability and interpretability. Consequently, Prasad teaches away from replacing, and it would not be obvious to replace, Prasad's rules with precategorized documents in combination of Pirolli's and Prasad's devices.

(b) ***Pirolli Teaches Away From the Claimed Invention***

In subsections "(i)" - "(iv)" below, various manners in which Pirolli teaches away from the claimed invention are discussed.

(i) *Categories in Pirolli Either do Not Rely on Similarity or Do Not have Documents Established as Belonging to Them*

A proposed modification cannot change the manner of operation in which the original device was intended to function (See MPEP 2143.01, p. 2100-127, the right column, entitled, "THE PROPOSED MODIFICATION CANNOT CHANGE THE PRINCIPLE OF OPERATION OF A REFERENCE," which cites *In re Ratti*, 270 F.2d 810, 123 USPQ 349

(CCPA 1959)). It logically follows that modifications that are incompatible with the manner in which a device works are not obvious.

Within the passage cited above, claims 1 and 34 recite "each training set is associated with a category and includes training documents that have been classified as belonging to said associated category...," which is also not taught by the combination of *Pirolli* and *Prasad*. In contrast to claims 1 and 34, *Pirolli* teaches that documents are categorized into functional categories that are (as recited by col. 8, lines 34 - 36),

> designed by someone (application designer, webmaster, end user), in contrast to being automatically induced.

In *Pirolli*, a number of characteristics are used to classify documents. Only one of these characteristics is based on similarity between a document and a particular set of documents. Consequently, the use of a similarity matrix in combination with a training set, as recited in claims 1 and 34, is incompatible with or at least has no place in the determination of these other categories. Stated differently, regarding these other categories, there is no motivation to use a similarity matrix for categories in which similarity is not a factor.

Regarding the characteristic that is based on similarity (csim), Pirolli states "csim, [is] the textual similarity of the item to its children based upon previous SCA calculation (column 508)." *Pirolli* further teaches that textual similarity is used to determine whether a page belongs to the category of head page (e.g., home page) (col. 9, lines 14 – 24).

> For Head Nodes (classification criteria 601), being the first pages of a collection of documents with like content, it is expected that such pages will have high text similarity between itself and its children, and would have a high average depth of its children, and that it would be more likely to be an entry point based upon actual user navigation patterns.

Thus, at best, *Pirolli* teaches that textual similarity between a page and the children of the page is used to determine the correspondence between the page and the category of home page. However, the category of home page is not a category to which the set of children have been established as belonging to, and is not analogous to the different sources of documents to which search queries are directed by *Prasad*. The claims, on the other hand, require the feature of using similarity between a document and a particular set of documents that were established as belonging to a category to determine the correspondence between the document and the category.

(ii) *Differences in Types of Categories Mitigate Against Combining Pirolli and Prasad*

Additionally, the Examiner (at paper #11, the first paragraph of page 4, cited above) was apparently referring to statements in the response, such as

> In fact, *Pirolli* seems to teach against such a feature because of the types of functional categories it discloses. For example, head node is a category which includes documents in which text similarity between the documents in this category is of little relevance. Examples of a set of documents that could be established in this category are Yahoo's home page, Google's home page, and the USPTO home page. It would seem that text similarity between these pages and another page would have very little relevance to whether the other page is a home page.

The Examiner apparently agreed that these other categories are not those recited in the claims, and apparently was stating that he was not relying upon them in making the rejection. The Examiner apparently also agreed that the categories in *Prasad* and *Pirolli* are of a different nature. However, the differences in their nature does, in fact, mitigate against combining references, because it is not clear that the categories of one reference are even

considered categories in the other reference despite the common use of the word "category," and the Examiner has not shown this to be the case. Logically, one of ordinary skill in the art would have expected that using the methods for arriving at one type of category will not necessarily work well for finding the other type of category, because one set of categories (*Prasad*'s) is related to sources within which to search (e.g., Google's or Yahoo!'s databases) for documents, while the other relates to the relevance of web pages to a focus, which may be another web locality such as a web page (i.e., the likelihood that someone looking at one web page will want to look at another web page and examples of such categories are head node or home page). Logically, unlike using a training set or sample set for finding common characteristics of documents in a source of documents, *Pirolli* teaches that deciding on whether a page is a head node or a home page is best done by a human.

(iii) *Pirolli's Global Rules, Requiring Human Intervention to Arrive at the Rules, Discourage use of Prasad's Automated Comparison to Documents to Arrive at Rules*

As stated in MPEP 2161, p. 2100-157, "Furthermore, '[k]nown disadvantages in old devices which would naturally discourage search for new inventions may be taken into account in determining obviousness.' *United States v. Adams,* 383 U.S. 39, 52, 148 USPQ 479, 484 (1966)."

Pirolli is concerned that (as recited in column 1, lines 25-28)

> Hypertext structures primarily affords information seeking by the sluggish process of browsing from one document to another along hypertext links. This sluggishness can be at least partly attributed to three sources of inefficiency in the basic process.

In other words, *Pirolli* is addressing the problem of the "sluggishness" associated with prior art searching techniques. *Pirolli* attributes the sluggishness in part to (as recited in column 1, lines 31-36),

> important information about the kinds of documents and content contained in the total collection cannot be immediately and simultaneously obtained by the user in order to assess the global nature of the collection or to aid in decisions about what documents to pursue.

In other words, *Pirolli* is attempting to find global generalizations about a collection, which at the time of the invention apparently one would not have expected to be able to satisfactorily extract from textual relationships. Consequently, the reason *Pirolli* et al. like the use of rules (rather than the textual similarity between a document and a training set with which the Examiner would like to modify *Pirolli*) is

> Based on category membership, a user may ***quickly*** predict the functionality of an element. For instance, in the everyday world, identifying something as a "chair" enables the quick prediction that an object can be sat on... (emphasis added, column 8, lines 53-55).

In other words, an important point being made here is that, for example, a reference about a chair may not mention anything about sitting, but by using rules one can nonetheless quickly make an association between the chair and sitting. Similarly, using rules one can make an association between a document and how to categorize it, even though the document may not explicitly mention anything about many of its attributes.

However, one of ordinary skill in the art would have expected that such an advantage in using rules "designed by someone (application designer, webmaster, end user), in contrast to being automatically induced" would be lost were one to use a bunch of training document to establish the rules because the rules established from training are unlikely to include concepts that are not explicitly in the training documents and because using training

documents increases the time to establish the rule. In other words the disadvantages of *Prasad*'s training set suggested by *Pirolli* would have discouraged such a combination. Therefore, one of ordinary skill in the art would be inclined not slow down the categorizing process by using the more limited rules derived from training documents of *Prasad* (that *Pirolli* would presumably have referred to as non-global rules). In this sense the *Prasad*'s use of the training documents runs contrary to at least one of the principals upon which *Pirolli* et al. are relying, which is not permitted in a rejection under 35 U.S.C. § 103. Similarly, the advantages of human derived heuristic rules over non-Global textual based rules taught by *Pirolli* would naturally discourage the use of the textual based rules automatically derived from the training set of *Prasad.*

(iv) *Added Effort in Applying Method of Prasad to Priolli as Compared to Applying Prasad in General is a Further Deterrent to Combining Prasad and Pirolli*

Further, the claims require that the training documents be already categorized into the categories. In *Prasad*, it would appear that the training documents happen to already be in the sources before the search began with no effort on the part the developer to categorize the documents. While the claims do not necessarily require effort or a pre-categorization step on the part of a developer, in the modification proposed by the Examiner, the effort of pre-classification typically required in finding training documents for categorizing (which is not necessary in *Prasad* because *Prasad* is deciding on which categories to use for sources), would have deterred one of ordinary skill from using training documents when categorizing, and would have caused one of ordinary skill in the art to think of these two activities as unrelated distinct processes.

(c) *With Hindsight Removed, One Would Not Have Arrived at the Modification Proposed by the Examiner*

As stated in MPEP 2141.01, " 'It is difficult but necessary that the decision maker forget what he or she has been taught . . . about the claimed invention and cast the mind back to the time the invention was made (often as here many years), to occupy the mind of one skilled in the art who is presented only with the references, and who is normally guided by the then-accepted wisdom in the art.' *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 USPQ 303, 313 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984)."

The main point of *Prasad* is about how to choose sources from which to search, while the main point of *Pirolli* is to use activation pumping on a set of web pages within a set of sources to determine their relevance to a focus. Thus, even arguendo were the above negative teachings not a deterrent to one of ordinary skill to have combined *Prasad* and *Pirolli*, following just the suggestions in *Prasad* and *Pirolli* without the hindsight benefit of the claims as a guide, the modification applied to *Pirolli* by one of ordinary skill would have been to use *Prasad*'s training documents to decide on which source to take the documents from and not in categorizing, ranking, or pumping activation to the documents later found within those sources. It would seem unlikely that one of ordinary skill in the art would look to a reference on where to search, to solve a problem about how to categorize search results *Prasad*.

(d) ***Prasad and Pirolli are From Different Fields of Endeavor***

As stated in MPEP 2141.01(a), p. 2100-117, "The examiner must determine what is 'analogous prior art' for the purpose of analyzing the obviousness of the subject matter at issue. 'In order to rely on a reference as a basis for rejection of an applicant's invention, the reference must either be in the field of applicant's endeavor or, if not, then be reasonably pertinent to the particular problem with which the inventor was concerned.' *In re Oetiker*, 977 F.2d 1443, 1446, 24 USPQ2d 1443, 1445 (Fed. Cir. 1992). See also *In re Deminski*, 796 F.2d 436, 230 USPQ 313 (Fed. Cir. 1986); *In re Clay*, 966 F.2d 656, 659, 23 USPQ2d 1058, 1060-61 (Fed. Cir. 1992). Logically, to combine two references they must be from the same field of endeavor. Otherwise, one of ordinary skill in the art would not be aware of the second reference or any motivations that might otherwise have been taught.

Prasad is attempting to determine from which source to retrieve documents, while *Pirolli* is attempting in-part to categorize documents found (which in a certain sense are contrasting or opposite features). Additionally, *Pirolli* is concerned with determining the relevance to a focus using pumping activation, which is not categorization *per se*. In this sense these two documents (*Prasad* and *Pirolli*) may not even be from related arts. Although both *Prasad* and *Pirolli* relate to documents and relate to categorization, they are no more related than the two SIMMS memories of *Wang Laboratories, Inc. v. Toshiba Corp.*, 993 F.2d 858, 26 USPQ2d 1767 (Fed. Cir. 1993) cited by MPEP 2141.01(a), p.2100-119, which relied on among other things, similar types of opposite or contrasting features such as one SIMM memory being compact and modular and the other being of varying sizes. Similarly, MPEP 2141.01(a) p.2100-118, which cites *In re Clay*, 966 F.2d 656, 23 USPQ2d 1058 (Fed.

Cir. 1992) and emphasizes the difference between "storage" and "extraction" as significant in determining a reference to be non-analogous. Storage in *In re Clay* is in at least some ways analogous to storage of search results, and extraction in *In re Clay* is in at least some ways analogous to the extraction of search results. Thus, the difference between storage and extraction is conceptually similar to the difference between categorizing search results and identifying sources of where to search.

The difference in the nature of the categories of *Pirolli* and *Prasad* (mentioned above) is further evidence that the two are not from the same art areas. Specifically, since the source where a document is found (e.g., whether to use Lexis', INSPEC's, or Dialog's databases, as discussed by *Prasad*) is not necessarily a useful category for the categorization of *Pirolli*. Since the Examiner has also not shown spreading activation of *Pirolli* to relate to the categorization of *Prasad* or of the claims, *Pirolli* appears to be from a different art area than *Prasad* and than the claimed invention.

C. CLAIM 8 IS NOT OBVIOUS BECAUSE THE EXAMINER HAS NOT THE CLAIMED SHOWN CLICK THROUGH BEHAVIOUR

Regarding claim 8, the Examiner cited column 10, lines 56-60, and FIG. 11, However, "the usage paths" and "flows of users through the locality" appear to be strengths of paths and flows between pages. Although these flows or paths may be the result of transfers from one page to another by clicking with a mouse, for example, they do not have to be. More importantly, however, the strength of paths or flows is not a comparison of the similarity of anything and therefore is not the similarity of click through behaviors related to two different documents.

D. CLAIM 9 IS NOT OBVIOUS BECAUSE THE EXAMINER HAS NOT SHOWN THE CLAIMED CLICK THROUGH BEHAVIOUR

Regarding claim 9, the Examiner cited column 11, lines 30-34, which states,

> Referring now to FIG. 13, for the matrix representation of usage path networks, an entry of an integer strength, s >=0, in column i row j, indicates the number of users that traversed from page i to page j.

The Examiner also cited column 7, lines 15-18, which states,

> From the set of paths, a vector that contains each page's frequency of requests is generated (i.e. a frequency vector), step 304, along with a path matrix containing the number of traversals from one page to another, step 305.

However, the strengths of paths, the strength of flows, and the number of traversals from one document to another are not in-and-of-themselves "similar patterns of user click behavior," because there is no comparison of behavior or other determination of similarity of behavior.

E. CLAIMS 17 AND 18 ARE NOT OBVIOUS BECAUSE THE EXAMINER HAS NOT SHOWN FORMING A GRAPH FROM TWO GRAPHS

Regarding claims 17 and 18, the Examiner cited column 10, lines 56-63, which state

> As outlined above, three kind of graphs, or networks, are used to represent strength of associations among Web pages: (1) the hypetext link topology of a Web locality, (2) inter-page text similarity, and (3) the usage paths, or flow of users through the locality. Each of these networks or graphs is represented by matrices in our spreading activation algorithm. That is, each row corresponds to a network node representing a Web page, and similarly each column corresponds to a network node representing a Web page. If we index the 1, 2, . . . , N Web pages, there would be i=1, 2, . . . N columns and j=1, 2, . . . N rows for each matrix representing a graph network.

Regarding claim 17, the Examiner implied the unsupported assertion that this passage is a suggestion of taking a union of graphs, and then concluded (without any further proof) that it

would have been obvious to take a union of graphs. However, the above paragraph only discusses using graphs to "represent strength of association," which is used for "spreading activation." The word "union" never appears in *Pirolli*'s specification. Further, there is not even a teaching or suggestion in column 10, lines 56-63, of forming the graph from two graphs, and therefore there is not any corresponding teaching or disclosure of forming the union of two graphs.

F. CLAIM 18 IS NOT OBVIOUS BECAUSE THE EXMINER HAS NOT SHOWN AN INTERSECTION OF TWO GRAPHS

Similarly, regarding claim 18, the above-cited paragraph never discusses taking the intersection of two graphs. Although the Examiner cites column 11, lines 1-34, and points out that the strength of association can be zero, there is no teaching in column 11, lines 1-34 (or column 10, lines 56-63), that the graph representing the strength of association was formed from a combination of two graphs. Consequently, column 10, lines 56-63, and column 11, lines 1-34, cannot teach or suggest taking the intersection of two graphs.

G. CLAIM 20 IS NOT OBVIOUS OVER BENGIO BECAUSE THE EXAMINER HAS NOT PROVIDED A PROPER MOTIVATION TO COMBINE AND BENGIO IS NONANALOGOUS ART

The Examiner rejected Claim 20 as unpatenable, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of U.S. Patent No. 6,128,606, herein *Bengio*.

Regarding claim 20, as a motivation to combine references, the Examiner stated (paper #9)

> However, Bengio, in disclosing an invention 'directed to the problem of developing a modular building block for complex processes that can input and output data in a wide variety of forms, but when interconnected with other similar modular building blocks can easily trained" (Bengio, col. 2, lines 45-49), teaches "training a network of these modules by back-propagating gradients through the network to determine a minimum of the global objective function." (Bengio, col. 2, lines 57-60.) Because claim 20 is directed to a similar invention, it would have been obvious to one of ordinary skill in the art to have combined Pirolli, Prasad, and Bengio to implement the optimization of an objective function.

However, the similarity of the contents of claim 20 and *Bengio* at best relates to whether *Bengio* is analogous art to claim 20. Although a prerequisite to *Bengio* being a proper reference under 35 U.S.C. § 103 is that *Bengio* must be analogous art, similarity of subject matter between claim 20 and *Bengio* is not a motivation for modifying *Prasad* or *Pirolli* by including a feature of *Bengio*. (Cf. MPEP 2143.01, p.2100-126, under the title "**FACT THAT REFERENCES CAN BE COMBINED OR MODIFIED IS NOT SUFFICIENT TO ESTABLISH *PRIMA FACIE* OBVIOUSNESS**," which states, "The mere fact that references can be combined or modified does not render the resultant combination obvious unless the prior art also suggests the desirability of the combination. *In re Mills*, 916 F.2d 680, 16 USPQ2d 1430 (Fed. Cir. 1990)." Also see MPEP 2143.01, p. 2100-125, for example, regarding the need to provide a motivation to combine references)

In fact, however, it is also not clear whether *Bengio* is analogous art, because *Bengio* state (in the last two sentences of the abstract)

> A complete check reading system based on these concept is described. The system uses convolutional neural network character recognizers, combined with global training techniques to provides record accuracy on business and personal checks.

Thus, *Bengio* is a check reading system, which is quite different than a system for predicting relevance of documents or a system for selecting a source for retrieving documents.

Bengio states (at column 1, lines 9-11)

> The present invention relates generally to modular networks and processes, and more particularly to a modular process in which each module receives data and outputs data that is structured as graphs.

Similarly, column 2, lines 45-47, cited by the Examiner, state

> The present invention is therefore directed to the problem of developing a ***modular building block*** for complex processes that can input and output data in a wide variety of forms, but when interconnected with other similar modular building blocks can be easily trained (emphasis added).

The Examiner has not shown that either of *Prasad* or *Pirolli* relate to one of the "modular networks and processes" of *Bengio* or that such modular building blocks are consistent with the teachings of *Prasad* and *Pirolli*, because including these modular building blocks within *Prasad* or *Pirolli* would appear to require that *Prasad* or *Pirolli*'s systems be completely rebuilt differently.

Although *Bengio* gives several motivations for the use of their system, the Examiner has not shown how any of these motivations are relevant to *Prasad* or *Pirolli*. *Bengio* is troubled by the problem of (column 1, line 55) "creating the ***intermediate data*** on which the module is to learn" (emphasis added). In contrast, the Examiner has not shown any discussion of "intermediate data" in *Prasad* or *Pirolli*. *Bengio* is concerned with (column 2, lines 3 and 4) "The limited flexibility of fixed–size vectors" and dealing with (column 2, line 14) "variable length sequence vectors," which the Examiner has also not shown to be relevant to either *Prasad* or *Pirolli*. *Bengio* state (column 1, lines 37-54)

> For example, a character recognition module can be trained to recognize well-formed individual characters. However, the role of the recognizer in the context of the entire

> system is usually quite different than simply recognizing the characters. Very often, character recognizers are expected to also reject badly segmented characters and other non-characters..... Merely training the character recognizer module to minimize its classification error on individual characters will not minimize the global objective function. Ideally it is desirable to find a good minimum of the global objective function with respect to all of the parameters in the system.

In other words, (in addition to the differences between character recognition and document retrieval or source identification) *Bengio* is concerned that optimizing the performance of an individual module without regard for how it interacts with the rest of the system may result in suboptimal performance for the system as a whole when the module is integrated into the entire system. The Examiner has not shown this concern to be relevant to *Pirolli* or *Prasad.*

Further, as cited by the Examiner (the Examiner actually only cited column 2, lines 57-60), *Bengio* also state (column 2, lines 52-60),

> The present invention solves this problem ***by using a graph transformer as a basic modular building block***, by using differentiable functions in each module to produce numerical data attached to an output graph from numerical data attached to an input graph and from any tunable parameters within the module, and by training a network of these modules by back-propagating gradients through the network to determine a minimum of the global objective function (emphasis added).

In other words, *Bengio* view their contribution ***not*** as the use of an optimization of an objective function mentioned in lines 57-60 (cited by the Examiner), but the use of a "graph transformer as a basic modular building block." Logically, the global nature of the optimization of *Bengio* links the optimization of the individual components. Consequently, if the system being modified by *Bengio* did not already include a ***global*** optimization of an objective function, the optimization of the individual modules would not be linked, and merely optimizing the individual modules may very well optimize the performance of the entire system. Thus, the use of an optimization of an objective function (in column 1, lines 37-54) is the source of the problem being solved by *Bengio* (and not the solution) and the

global optimization is assumed to be the background in which *Bengio* employ their system. It follows that systems such as *Pirolli*, lacking any optimization of a global function, have no need for the device of *Bengio*. The Examiner has also not shown *Bengio* to give a motivation for using an optimization of an objective function in other contexts. Thus, it would appear that one of ordinary skill in the art would not have thought of applying the globally optimizable modules of *Bengio* to a system (such as that of *Pirolli* or *Prasad*) that (1) did not already require a global optimization of an objective function and therefore (2) does not require a global optimization when each module is added.

H. CLAIMS 20-25 BECAUSE THERE IS NO MOTIVATION TO COMBINE, SINCE IT DOES NOT MAKE SENSE TO PERFORM SPREADING ACTIVATION VIA OPTIMIZING AN OBJECTIVE FUNCTION AS REQUIRED FOLLOWING THE EXAMINER"S LINE OF REASONING

The Examiner rejected Claim 20 as unpatentable, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of U.S. Patent No. 6,128,606, herein *Bengio*. The Examiner also rejected Claims 21-25 as unpatentable, under 35 U.S.C. § 103(a), over *Pirolli* and *Prasad* further in view of *Chakrabarti*.

Regarding claims 20-25, the Examiner associated the use of spreading activation to define degrees of predicted relevance to meet the recitation of "extracting similarity information from the similarity matrix" of claim 19. Claims 20-25 directly or indirectly depend upon claim 19, and claims 20-22 further define the extraction of the similarity information as being performed by optimizing or approximately optimizing an objective function. Although claims 23-25 do not recite the optimization of an objective function, the

Examiner relied on aspects of *Chakrabarti*'s optimization of an objective function in rejecting claims 23-25.

However, spreading activation and optimizing an objective function are quite different. The spreading activation process involves repeating the step of passing a "token" or a signal from one node to all other nodes that are linked to that node. The token or signal is typically not necessarily a real token or signal, but a numerical value. An entry node is chosen to start the process. As the signal is passed along the "arcs" (using *Pirolli*'s terminology), the signal is attenuated. The stronger the arc connecting the nodes the less the signal is attenuated (*Pirolli* therefore refers to the arcs having a "capacity"). To simplify the computation, a stopping criterion is often chosen. As examples of stopping criteria, often once the signal is below a certain strength (e.g., .01 of its initial value) or after the original signal has propagated a certain number of arcs (or after a certain number of "iterations"), the signal is no longer propagated along the arcs. Each time a signal is passed to a node its strength is added to a value associated with the node. From these values, after the process ends, conclusions can be drawn about the strength of association of the various nodes to the entry node.

In contrast, the optimization of an objective function referred to by *Chakrabarti* or *Bengio* is typically a minimization or maximization of a multivariable function performed by adjusting the values of the variables. The Applicants respectfully submit that it does not make sense to perform spreading activation via an optimization of an objective function, contrary to the implications of the Examiner's rejections (Cf. MPEP 2143.01 and *In re Ratti,* cited above). For example, regarding claim 22, growth transformations have absolutely no place in the spreading activation process.

I. REMAINING DEPENDANT CLAIMS

The pending claims not discussed so far are dependant claims that depend on an independent claim that is discussed above. Because each of these dependant claims include the limitations of claims upon which they depend, the dependant claims are patentable for at least those reasons the claims upon which the dependant claims depend are patentable. Reconsideration these dependant claims and allowance of the dependant claims is respectfully requested.

IX. CONCLUSION AND PRAYER FOR RELIEF

The rejections of the final Office Action under 35 U.S.C. § 103(a) lack the requisite factual and legal basis. The applied references, *Prasad, Pirolli, Hoffert, Bengio, and Chakrabarti*, do not disclose or suggest the numerous features of the rejected claims for the specific reasons discussed above.

Appellants therefore respectfully submit that the rejections under 35 U.S.C. § 103(a) are incorrect and respectfully solicit the Board to **reverse** each of the imposed rejections under 35 U.S.C. § 103(a) and to remand the case to the Examiner for further proceedings.

Respectfully submitted,

HICKMAN PALERMO TRUONG & BECKER LLP

Dated: Oct 18, 2003

David Lewis
Agents for Applicants/Appellant
Reg. No. 33,101

1600 Willow Street
San Jose, CA 95125
Telephone No.: (408) 414-1080 ext.213
Facsimile No.: (408) 414-1076

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Mail Stop Appeal Brief - Patents, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

on 10/20/03 (Date) by (Signature)

APPENDIX
The Pending Claims

CLAIMS

What is claimed is:

1. A method of categorizing a plurality of new electronic documents into a set of categories, comprising the steps of:

establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;

determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories by determining similarity between said each document and the training documents that belong to the training set of said category; and

wherein the step of determining similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity.

2. A method as recited in Claim 1, wherein the measures of document similarity include hyperlink similarity.

3. A method as recited in Claim 2, in which two documents among the plurality of documents are considered similar to each other when there is a link from one to the other, or when the two documents link to, or are linked to by, a set of other associated documents.

4. A method as recited in Claim 3, in which certain hyperlinks have greater or lesser similarity weight than other hyperlinks, based on other features of the links or their source or destination documents.

5. A method as recited in Claim 1, wherein the measures of document similarity include a similarity of text of the documents.

6. A method as recited in Claim 5, wherein two documents are considered similar based on a comparison of word vectors derived from the text of each of the two documents.

7. A method as recited in Claim 5, wherein text similarity is determined in part based upon weight values assigned to words of the text, and wherein certain words have greater or lesser weight than other words.

8. A method as recited in Claim 1, wherein the measures of document similarity include user click-through similarity.

9. A method as recited in Claim 8, wherein two documents are considered similar based on user click-through similarity when the documents are associated with similar patterns of user click behavior, selected from among frequency of clicks, click context, duration of viewing, proximity in time to other clicks, or proximity in context to other clicks.

10. A method as recited in Claim 1, wherein the measures of document similarity are derived from patterns detected in user viewing of the documents.

11. A method as recited in Claim 10, wherein the user viewing information is monitored by a web caching system and stored in a log.

12. A method as recited in Claim 10, wherein two documents are considered similar based on patterns of user viewing behavior, including frequency of viewing, viewing context, duration of viewing, proximity in time to other documents viewed by the same user, or similarity of patterns of viewing by all users.

13. A method as recited in Claim 1, wherein the measures of document similarity include URL similarity.

14. A method as recited in Claim 13, wherein two documents are considered similar if a URL of each document contains similar URL sub-components.

15. A method as recited in Claim 1, wherein the measures of document similarity include multimedia similarity.

16. A method as recited in Claim 15, wherein two documents are considered similar based on features derived from multimedia components linked to or contained by the documents.

17. A method as recited in Claim 1, wherein the combination of two or more measures of document similarity is achieved by taking the union of each of a plurality of graphs, each graph describing one of the measures of document similarity, to compute a combined graph that describes the combined document similarity.

18. A method as recited in Claim 1, wherein the combination of two or more measures of document similarity is achieved by taking the intersection of each of a plurality of graphs, each graph describing one of the measures of document similarity, to compute a combined graph that describes the combined document similarity.

19. (Amended) A method as recited in Claim 1, further comprising the step of extracting similarity information from the similarity matrix to obtain new documents supported by the set of training documents for each category.

20. (Amended) A method as recited in Claim 19, wherein the similarity information is obtained by optimizing an objective function.

21. (Amended) A method as recited in Claim 19, wherein the similarity information is obtained by only approximately optimizing an objective function.

22. A method as recited in Claim 21, wherein approximately optimizing the objective function comprises repeated application of a growth transformation.

23. A method as recited in Claim 19, further comprising the step of creating and storing a second matrix that represents an interim score for each document in each category.

24. A method as recited in Claim 19, further comprising the steps of, periodically as the matrix is being computed, normalizing rows of the matrix by normalizing within each document, across all categories, whereby the score for one document in a particular category will depend on the scores for that document in all other categories.

25. A method as recited in Claim 19, further comprising the steps of, periodically as the matrix is being computed, normalizing columns of the matrix by normalizing within each category, across all documents, whereby the score for one document in a particular category depends on the scores for all other documents in that category.

26. A method as recited in Claim 1, in which the categories come from a manually defined taxonomy.

27. A method as recited in Claim 1, wherein the categories are derived from logs of user queries.

28. A method as recited in Claim 1, further comprising the steps of creating and storing a second matrix using columns representing documents and rows representing user sessions, and wherein values of elements of the second matrix represent interest in a document shown by a particular user in a particular session.

29. A method as recited in Claim 1, further comprising the steps of creating and storing a matrix using columns representing user sessions and rows representing documents, and wherein values of elements of the second matrix represent interest in a document shown by a particular user in a particular session.

30. A method as recited in Claim 28, wherein the element values are computed as a function of a time that a user has spent viewing a document associated with each element.

31. A method as recited in Claim 28, further comprising the steps of creating and storing a second matrix representing a Similarity between pairs of documents i and j, wherein the second matrix is derived by comparing pairs of column vectors or row vectors, respectively i and j of the first matrix.

32. A method as recited in Claim 28, further comprising the steps of creating and storing a second matrix representing a Similarity between pairs of documents i and j, by finding pairs of documents i and j which have high interest values for a particular user in a particular session or period of time.

33. The method recited in Claim 1, further comprising the steps of:
identifying a category of a classification taxonomy of the hypertext system in which a first electronic document is presently classified; and
if a second electronic document is found to be highly Similar, storing information that classifies the second electronic document into the category.

34. A computer-readable medium carrying one or more sequences of instructions, wherein execution of the one or more sequences of instructions by one or more processors causes the one or more processors to perform the steps of:
establishing a plurality of training sets, wherein each training set is associated with a category and includes training documents that have been classified as belonging to said associated category;
determining how strongly each document of said plurality of documents corresponds to each of said plurality of categories by determining similarity between said each document and the documents that belong to the training set of said category; and

wherein the step of determining similarity is performed using a matrix representing document similarity that is derived by combining two or more measures of document similarity.

35. (Cancelled)

36. (Cancelled)

37. (Cancelled)

AF 2700

PTO/SB/21 (08-00)
Approved for use through 10/31/02. OMB 0651-0031
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays avalid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	09/333,121
Filing Date	June 14, 1999
First Named Inventor	Michael E. Palmer
Group Art Unit	2176
Examiner Name	Charles A. Bieneman
Attorney Docket Number	50269-0026

ENCLOSURES ***(check all that apply)***

- [X] Fee Transmittal Form
  - [X] Fee Attached
- [ ] Amendment / Response
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Response to Missing Parts/ Incomplete Application
  - [ ] Response to Missing Parts under 37 CFR 1.52 or 1.53
- [ ] Assignment Papers *(for an Application)*
- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition To Convert To a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, number of CD(s)_____
- [ ] After Allowance Communication to Group
- [ ] Appeal Communication to Board of Appeals and Interferences
- [X] Appeal Communication to Group *(Appeal Notice, Brief, Reply Brief)*
- [ ] Proprietary Information
- [ ] Status Letter
- [ ] Other Enclosure(s) *(please identify below):*

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm or Individual name	Hickman Palermo Truong & Becker LLP David Lewis, Patent Agent, Reg. No. 33,101
Signature	David Lewis Reg. No. 33,101
Date	October 20, 2003

CERTIFICATE OF MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class: mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on this date: October 20, 2003

Type or printed name	Teresa Austin		
Signature	Teresa Austin	Date	October 20, 2003

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time your are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

PTO/SB/17 (12/99)
Approved for use through 09/30/2000. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2003

Patent fees are subject to annual revision,
Small Entity payments must be supported by a small entity statement, otherwise large entity fees must be paid. See Forms PTO/SB/09-12.
See 37 C.F.R. §§ 1.27 AND 1.28

TOTAL AMOUNT OF PAYMENT ($) 330.00

Complete if Known	
Application Number	09/331,121
Filing Date	June 14, 1999
First Named Inventor	Michael E. Palmer
Examiner Name	Charles A. Bieneman
Group/Art Unit	2176
Attorney Docket No.	50269-0026

METHOD OF PAYMENT (check one)

1. [X] Throughout the pendency of this application, please charge any additional fees, including any required extension of time fees, and credit all overpayments to deposit account 50-1302. A duplicate of this sheet is enclosed.

Deposit Account Number: 50-1302

Deposit Account Name: Hickman Palermo Truong & Becker, LLP

2. [X] Payment Enclosed:
[X] Check [ ] Money Order [ ] Other

3. [ ] Applicant(s) is entitled to small entity status. See 37 CFR 1.27

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	

SUBTOTAL (1) ($) 0.00

2. EXTRA CLAIM FEES

		Extra Claims		Fee from Below		Fee Paid
Total Claims	-20**=	0	X	18.00	=	0.00
Independent Claims	- 3**=	0	X	84.00	=	0.00
Multiple Dependent					=	

**or number previously paid, if greater; For Reissues, see below

Large Entity Fee Code	Large Entity Fee ($)	Small Entity Fee Code	Small Entity Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	**Reissue independent claims over original patent
1205	18	2205	9	**Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($) 0.00

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity Fee Code	Large Entity Fee $	Small Entity Fee Code	Small Entity Fee $	Fee Description	Fee Paid
1051	130	2051	65	Surcharge – late filing fee or oath	
1052	50	2052	25	Surcharge – late provisional filing fee or cover sheet.	
1053	130	1053	130	Non-English specification	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1,840*	1805	1,840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1,450	2254	725	Extension for reply within fourth month	
1255	1,970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	330	2402	160	Filing a brief in support of an appeal	330.00
1403	280	2403	140	Request for oral hearing	
1451	1,510	1451	1,510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive – unavoidable	
1453	1,300	2453	650	Petition to revive – unintentional	
1501	1,300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Petitions related to provisional applications	
1806	180	1806	180	Submission of information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR § 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR § 1.129(b))	

Other fee (specify) _____

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid SUBTOTAL (3) ($) 330.00

SUBMITTED BY

Complete (if applicable)

Name (Print/Type)	David Lewis	Registration No. (Attorney/Agent)	33,101	Telephone	(408) 414-1080
Signature				Date	October 20, 2003

WARNING: **Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.**

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Alexandria, VA 22313-1450.
DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.